ORIGINAL ARTICLE

Cross-National Associations Between Gender and Mental Disorders in the World Health Organization World Mental Health Surveys

Soraya Seedat, PhD; Kate Margaret Scott, PhD; Matthias C. Angermeyer, PhD; Patricia Berglund, MBA; Evelyn J. Bromet, PhD; Traolach S. Brugha, MD (NUI), FRCPsych; Koen Demyttenaere, MD, PhD; Giovanni de Girolamo, MD; Josep Maria Haro, MD, MPH, PhD; Robert Jin, MA; Elie G. Karam, MD; Viviane Kovess-Masfety, MD, PhD; Daphna Levinson, PhD; Maria Elena Medina Mora, PhD; Yutaka Ono, MD, PhD; Johan Ormel, PhD; Beth-Ellen Pennell, MA; Jose Posada-Villa, MD; Nancy A. Sampson, BA; David Williams, PhD, MPH; Ronald C. Kessler, PhD

Context: Gender differences in mental disorders, including more anxiety and mood disorders among women and more externalizing disorders among men, are found consistently in epidemiological surveys. The gender roles hypothesis suggests that these differences narrow as the roles of women and men become more equal.

Objectives: To study time-space (cohort-country) variation in gender differences in lifetime *DSM-IV* mental disorders across cohorts in 15 countries in the World Health Organization World Mental Health Survey Initiative and to determine if this variation is significantly related to time-space variation in female gender role traditionality as measured by aggregate patterns of female education, employment, marital timing, and use of birth control.

Design: Face-to-face household surveys.

Setting: Africa, the Americas, Asia, Europe, the Middle East, and the Pacific.

Participants: Community-dwelling adults (N=72 933).

Main Outcome Measures: The World Health Organization Composite International Diagnostic Interview assessed lifetime prevalence and age at onset of 18 *DSM-IV* anxiety, mood, externalizing, and substance disorders. Survival analyses estimated time-space variation in female to male odds ratios of these disorders across cohorts defined by the following age ranges: 18 to 34, 35 to 49, 50 to 64, and 65 years and older. Structural equation analysis examined predictive effects of variation in gender role traditionality on these odds ratios.

Results: In all cohorts and countries, women had more anxiety and mood disorders than men, and men had more externalizing and substance disorders than women. Although gender differences were generally consistent across cohorts, significant narrowing was found in recent cohorts for major depressive disorder and substance disorders. This narrowing was significantly related to temporal (major depressive disorder) and spatial (substance disorders) variation in gender role traditionality.

Conclusions: While gender differences in most lifetime mental disorders were fairly stable over the time-space units studied, substantial intercohort narrowing of differences in major depression was found to be related to changes in the traditionality of female gender roles. Additional research is needed to understand why this temporal narrowing was confined to major depression.

Arch Gen Psychiatry. 2009;66(7):785-795

Author Affiliations are listed at the end of this article.

EPIDEMIOLOGICAL SURVEYS have consistently documented significantly higher rates of anxiety and mood disorders among women than men[1,2] and significantly higher rates of externalizing and substance use disorders among men than women.[3-5] Although a number of biologic, psychosocial, and biopsychosocial hypotheses have been proposed to account for these patterns,[6-8] evidence that gender differences in depression[9,10] and substance use[11-13] have narrowed in a number of countries has led to a special interest in the gender roles hypothesis. This hypothesis asserts that gender differences in the prevalence of mental disorders are due to differences in the typical stressors, coping resources, and opportunity structures for expressing psychologic distress made available differentially to women and men in different countries at different points throughout history.[14,15] Consistent with this hypothesis, evidence of decreasing gender differences in depression and substance use has been found largely in countries in which the roles of women have improved in terms of opportunities for employment, access to birth control, and other indicators of increasing gen-

der equality, while trend studies in countries in which gender roles have been more static[11,16] or during historical periods when gender role changes have been small[17] have failed to document a reduction in gender differences in depression or substance use.

Most research aimed at investigating the gender roles hypothesis has focused on individual-level variation in roles in a single country at a single time.[18-20] This approach is limited in 3 ways. First, selection bias for roles owing to preexisting mental illness (eg, women with agoraphobia having a higher probability than other women of becoming homemakers rather than seeking employment outside the home) confounds attempts to evaluate the causal effects of gender roles. Second, gender differences are largely confined to differences in lifetime risk, with much less evidence for gender differences in recent prevalence among lifetime cases.[21] This means that investigation of the determinants of gender differences should focus on lifetime first onset of disease rather than on the recent prevalence that has been the focus of most studies. Third, because the gender roles hypothesis is one about the effects of social context, a rigorous test of the hypothesis requires an analysis of societal-level, time-space variation rather than the individual-level variation that has been the focus of most studies.

A small number of cross-national comparative studies have examined spatial variation in gender differences in depression[22] and alcohol abuse[13] at 1 point or, more rarely, at 2 points.[11] Although these studies raised the possibility that gender roles might be associated with variation in the magnitude of gender differences in these outcomes, they were unable to test this hypothesis owing to the small number of cross-sectional country-level observations included in the analyses. Our article provides a more direct test of the gender roles hypothesis by analyzing community epidemiological data collected from respondents surveyed in 15 countries as part of the World Health Organization World Mental Health (WMH) Survey Initiative.[21] Previous cross-national comparisons of gender differences in mental illness focused on cross-sectional differences. We, in comparison, use retrospective reports obtained in the WMH surveys about lifetime occurrence and age at onset of mental disorders in different birth cohorts to study time-space variation in lifetime risk. Specifically, we examine both variation across cohorts within a single country (temporal variation) and variation across countries within a single cohort (spacial variation) in lifetime risk of mental disorders as a function of time-space variation in the traditionality of gender roles. We focused on lifetime risk rather than recent prevalence, though accuracy of reporting is doubtlessly better for recent episodes than lifetime occurrence, to address the fact that gender differences in lifetime risk are much more robust than those in current prevalence among lifetime cases.

METHODS

STUDY SAMPLE

World Mental Health surveys were carried out in samples of adults (age ≥18 years) in 5 countries classified by the World Bank[23] as developing (Colombia, Lebanon, Mexico, South Africa, and Ukraine) and 10 classified as developed (Belgium, France, Germany, Israel, Italy, Japan, the Netherlands, New Zealand, Spain, and the United States). The total sample size was 72 933 individuals. Individual country samples ranged from 2372 (the Netherlands) to 12 790 (New Zealand) (**Table 1**). The weighted average response rate was 71.2%. Country-specific response rates ranged from 45.9% (France) to 87.7% (Colombia). All surveys were based on probability household samples either regionally representative (Colombia, Japan, and Mexico) or nationally representative (all other countries). Survey sample characteristics are described in more detail elsewhere.[24]

All interviews were conducted face-to-face by trained lay interviewers. Each interview had 2 parts. All respondents completed part I, which contained assessments of core mental disorders; all part I respondents who met criteria for any core disorder plus a probability subsample of approximately 25% of other part I respondents were administered part II. Part II assessed correlates, service use, and disorders of secondary interest. The part II data, used in the current article, were weighted to adjust for oversampling part I respondents with mental disorders and for differential probabilities of selection within households (owing to only 1 household member, and in some cases 2, being surveyed in each household no matter the number of adults residing there) and to match sample distributions to population sociodemographic distributions. Standardized interviewer training procedures, translation and back-translation procedures, and quality-control procedures were applied across all WMH countries to ensure comparability. Informed consent was obtained in all countries. These procedures are described in more detail elsewhere.[24,25]

DSM-IV DISORDERS

Mental disorders were assessed with version 3.0 of the World Health Organization Composite International Diagnostic Interview (CIDI),[26] a fully structured diagnostic interview. Translation, back-translation, and harmonization of the interview to local languages with the original English version of CIDI were carried out in each WMH country using World Health Organization guidelines.[27] Disorders were assessed using *DSM-IV* definitions.[28] Disorders assessed included mood disorders (major depressive disorder [MDD], dysthymic disorder, and bipolar disorder), anxiety disorders (panic disorder, generalized anxiety disorder, agoraphobia without panic disorder, social phobia, specific phobia, separation anxiety disorder, and posttraumatic stress disorder), externalizing disorders (attention-deficit/hyperactivity disorder, conduct disorder, intermittent explosive disorder, and oppositional defiant disorder), and substance disorders (alcohol and illicit drug abuse with or without dependence). The *DSM-IV* organic exclusion rules were used to make diagnoses.

Methodological evidence collected in clinical reappraisal studies show that diagnoses of anxiety, mood, and substance disorders based on CIDI have generally good concordance (25th-75th percentiles of area under the receiver operating characteristic curve equal to 0.71-0.81), with diagnoses based on blinded clinical reappraisal interviews.[29] No evaluations were made of test-retest reliability. The evidence regarding good concordance with clinical diagnoses is based on surveys that have been carried out in only a small number of countries. It is not clear that translations of the instrument in all countries yield data that would have the same good concordance with blinded clinical reappraisal interviews. In addition, the externalizing disorder diagnoses were not validated in the CIDI clinical reappraisal studies, because the clinical interview that was used as the gold standard in these studies did not assess externalizing disorders. However, a subsequent independent clinical calibration study documented good

Table 1. WMH Survey Sample Characteristics

Country	Survey	Sample Characteristics[a]	Survey Year	Age Range, y	Sample Size			Response Rate, %[c]
					Part I	Part II	Part II and Age ≤44 y[b]	
Developed Countries								
Belgium	ESEMeD	NR, stratified, multistage clustered probability sample of individuals in households from the national register of Belgium residents	2001-2002	≥18	2419	1043	486	50.6
France	ESEMeD	NR, stratified, multistage clustered sample of working telephone numbers merged with a reverse directory (for listed numbers). Initial recruitment was by telephone, with supplemental in-person recruitment in households with listed numbers	2001-2002	≥18	2894	1436	727	45.9
Germany	ESEMeD	NR, stratified, multistage clustered probability sample of individuals from community resident registries	2002-2003	≥18	3555	1323	621	57.8
Israel	NHS	Stratified, multistage clustered-area probability sample of individuals from a national resident registry	2002-2004	≥21	4859			72.6
Italy	ESEMeD	NR, stratified, multistage clustered probability sample of individuals from municipality resident registries	2001-2002	≥18	4712	1779	853	71.3
Japan	WMHJ 2002-2003	Unclustered 2-stage probability sample of individuals residing in households in 4 metropolitan areas (Fukiage, Kushikino, Nagasaki, and Okayama)	2002-2003	≥20	2436	887	282	56.4
The Netherlands	ESEMeD	NR, stratified, multistage clustered probability sample of individuals in households listed in municipal postal registries	2002-2003	≥18	2372	1094	516	56.4
New Zealand	NZMHS	NR, stratified, multistage clustered-area probability sample of household residents	2004-2005	≥18	12 790	7312	4119	73.3
Spain	ESEMeD	NR, stratified, multistage clustered-area probability sample of household residents	2001-2002	≥18	5473	2121	960	78.6
United States	NCS-R	NR, stratified multistage clustered-area probability sample of household residents	2002-2003	≥18	9282	5692	3197	70.9
Developing Countries								
Colombia	NSMH	Stratified multistage clustered-area probability sample of household residents in all urban areas of the country (approximately 73% of the total national population)	2003	18-65	4426	2381	1731	87.7
Lebanon	LEBANON	NR, stratified, multistage clustered-area probability sample of household residents	2002-2003	≥18	2857	1031	595	70.0
Mexico	M-NCS	Stratified multistage clustered-area probability sample of household residents in all urban areas of the country (approximately 75% of the total national population)	2001-2002	18-65	5782	2362	1736	76.6
South Africa	SASH	NR, stratified, multistage clustered-area probability sample of household residents	2003-2004	≥18	4351			87.1
Ukraine	CMDPSD	NR, stratified, multistage clustered-area probability sample of household residents	2002	≥18	4725	1720	541	78.3

Abbreviations: CMDPSD, Comorbid Mental Disorders During Periods of Social Disruption; ESEMeD, European Study of the Epidemiology of Mental Disorders; LEBANON, Lebanese Evaluation of the Burden of Ailments and Needs of the Nation; M-NCS, Mexico National Comorbidity Survey; NCS-R, National Comorbidity Survey Replication; NHS, National Health Survey; NR, nationally representative; NSMH, National Study of Mental Health; NZMHS, New Zealand Mental Health Survey; SASH, South Africa Health Survey; WMH, World Health Organization World Mental Health; WMHJ 2002-2003, World Mental Health Japan Survey.

[a] Most WMH surveys are based on stratified, multistage clustered-area probability household samples in which samples of areas equivalent to counties or municipalities in the United States were selected in the first stage followed by 1 or more subsequent stages of geographic sampling (eg, towns within counties, blocks within towns, and households within blocks) to arrive at a sample of households, from which a listing of household members was created; 1 or 2 people were selected from this listing to be interviewed. No substitution was allowed when the originally sampled resident could not be interviewed. These household samples were selected from census area data in all countries except for France (where telephone directories were used to select households) and the Netherlands (where postal registries were used to select households). Several WMH surveys (Belgium, Germany, and Italy) used municipal resident registries to select respondents without listing households. The Japanese sample is the only totally unclustered sample, with households randomly selected in each of the 4 sample areas and 1 random respondent selected in each sample household. Twelve of the 15 surveys are based on NR household samples, while 2 others are based on nationally representative household samples in urbanized areas (Colombia and Mexico).

[b] All countries with the exception of Ukraine (which was restricted to age ≤39 years) were restricted to age 44 years and younger.

[c] The response rate is calculated as the ratio of the number of households in which an interview was completed to the number of households originally sampled, excluding from the denominator households known not to be eligible either because of being vacant at the time of initial contact or because the residents could not speak the designated languages of the survey. The weighted average response rate is 71.2%.

concordance between diagnoses of adult attention-deficit/hyperactivity disorder based on CIDI and those based on blinded clinical reappraisal interviews.[30] Another problem exists with the diagnoses of substance dependence, which was assessed only among respondents who had a history of abuse. This means that cases of dependence without abuse are excluded. However, empirical studies in the United States have shown that the number of cases of dependence without a history of abuse is small and that their exclusion does not have a substantively meaningful effect on the estimated associations of predictors with the outcomes.[31-33] Nonetheless, because of this exclusion, we focus herein on abuse rather than dependence. Retrospective reports of age at onset were obtained with the CIDI using a series of questions designed to avoid the implausible response patterns obtained in reply to a simple question asking for recall of age at first episode of a focal disorder.[34]

GENDER ROLE TRADITIONALITY

Each respondent was classified as being in 1 of 4 birth cohorts defined by age at interview (18-34, 35-49, 50-64, and ≥65 years) to distinguish broad life-course stages (early adulthood, early midlife, late midlife, and old age, respectively). Four within-cohort indicators of female gender role traditionality were calculated in each of the 58 resulting time-space subsamples (4 cohorts in each of 15 countries, minus the 2 oldest cohorts in Colombia and Mexico owing to an upper age limit of 64 years in those 2 surveys). The 4 indicators were (1) the ratio of the proportion of women to men in the cohort who had labor force experience before age 35 years (extrapolated using survival analysis in the 18- to 34-year-old cohort and calculated directly in the older cohorts); (2) the ratio of the proportion of women to men in the cohort who achieved the median level of education found among workers in the upper quartile of the income distribution in the cohort; (3) the ratio of the median ages of marriage of women vs men in the cohort; and (4) the proportion of women in the cohort who used birth control pills or other medical forms of contraception before age 25 years (restricted to women aged 25-34 years in the 18- to 34-year-old cohort).

We make no claim that these indicators form an exhaustive set of defining characteristics of gender role traditionality or that the cutoffs used to construct the measures (eg, labor force participation by age 34 years rather than some other age we might have selected) are optimal. Rather, the indicators were constructed from the WMH survey data on an ad hoc basis to operationalize aspects of gender roles that we considered important based on our reading of the demographic literature on gender roles.[35-37]

Confirmatory factor analysis carried out at the level of the time-space unit (n=58) showed that our initial thinking in selecting the 4 indicators was correct in the sense that a strong single-factor structure was found among these 4 indicators, with factor loadings in the range of 0.59 to 0.91. (Detailed results showing the values of each gender role traditionality indicator for each time-space unit, the correlation matrix among the indicators, and factor loadings are available from the corresponding author on request.) This finding confirmed that the indicators are, in fact, strongly related and can be used to construct a composite measure that we interpret as a measure of gender role traditionality. This is the key predictor variable in the analysis we will describe.

Rather than use the ad hoc gender role traditionality measure described previously, it would have been preferable to obtain objective administrative data on country-level trends in indicators of gender role traditionality. Our attempts to obtain such data, though, were unsuccessful because of sparse historical data on these indicators in most countries. The Global Economic Forum collected country-level data of this sort to assess the socioeconomic-political positions of women in 58 countries in the year 2000,[38] but they, like us, were unable to obtain retrospective trend data. The goal of the Global Economic Forum was to create a baseline measure that could be used to track the United Nations Millennium Development Goals of gender equality in social, economic, and political functioning (http://www.un.org/millenniumgoals/). The Global Economic Forum report constructed a country-level gender empowerment measure (GEM) for this purpose, which summarized objective data on the economic opportunity and participation of women in each country along with data on female political empowerment, educational attainment, life expectancy, and access to health care (legal birth control and legal abortion). The GEM could be developed for only 58 countries because of missing data in the others. The GEM scores are unavailable for 2 WMH countries (Lebanon and Ukraine). Because the GEM measure was developed only for the year 2000, we could not use it to study within-country changes in gender equality over time. However, we were able to compare scores on the composite WMH gender role traditionality measure with GEM scores for the 13 WMH countries for which they were available to validate our survey-based measure against the gold standard GEM measure. The Pearson correlation between the 2 measures was found to be 0.78. This high correlation strongly suggests that our gender role traditionality measure validly assesses the traditionality of women's roles in the WMH countries.

STATISTICAL ANALYSIS

Gender differences in lifetime risk of each disorder were examined using discrete time survival analysis with person-years as the unit of analysis.[39] This is a method that takes into consideration age at onset of the disorder in examining predictors, making it possible to study the predictors of lifetime occurrence of the disorder among respondents who vary by age. Each year in the life of each respondent up to and including the age at onset of the focal disorder (or, in the case of respondents who never had the disorder, up to his or her age at interview) was treated as a separate observational record in this analysis, with the year at first onset coded 1 on a dichotomous outcome variable and earlier years coded 0. Years after first onset were excluded from the data file. Logistic regression was used to analyze these data, with gender (coded 1 for women, 0 for men), cohort (coded into the 4 categories as noted), and person-years (age at the time of the person-year observational record) included as predictors of first onset of the disorder. The logistic regression coefficients and their standard errors were exponentiated to create odds ratios (ORs) and 95% confidence intervals for ease of interpretation. Female to male (F:M) ORs are the main focus of attention.

A separate model was estimated for each *DSM-IV* and CIDI disorder separately in each country. These models were also estimated in a data file that combined observations across all countries. The cross-national models included 14 dummy predictor variables to distinguish among the 15 countries in addition to the other predictors. The basic models were elaborated to consider possible nonlinear effects of cohort and person-year (using polynomials and dummy variables to define ranges on these continuous variables) and to assess whether the gender difference in lifetime risks of the disorders varied by cohort, life-course phase, or country. Gender differences were for the most part consistent across the life course, so these results are not reported herein but are available on request. The models were then estimated a final time in the subsample of person-years in the age range of 1 to 34 years (up to the oldest age in the youngest grouped cohort subsample) to remove the association between cohort and age in the person-year data file.

In cases for which the survival analysis documented significant time-space variation in the gender difference for a particular outcome, structural equation models using the 58 time-space subsamples as the unit of analysis estimated the extent to which a latent measure of gender role traditionality (defined in terms of the 4 indicators described) could

Table 2. Associations of Gender With Lifetime Risk of *DSM-IV* Mental Disorders in the WMH Surveys (N=72 933)[a]

Mental Disorder	No. of Subjects[b]	All-Country F:M OR (95% CI)	Within-Country OR Variation				Gender × Country Interaction, χ^{2c}
			Range		%		
			Min	Max	Dominant Direction	$P \leq .05$ in Dominant Direction	
Mood disorders							
Major depressive disorder	15	1.9 (1.8-2.0)[d]	1.6	2.4	100.0	100.0	29.3[d]
Dysthymic disorder	10	1.9 (1.6-2.2)[d]	1.3	3.8	100.0	70.0	13.0
Bipolar disorder	6	0.9 (0.8-1.0)	0.6	1.1	83.3	20.0	8.1
Any mood disorder	15	1.8 (1.7-1.8)[d]	1.5	2.5	100.0	100.0	47.7[d]
Anxiety disorders							
Panic disorder	12	1.9 (1.7-2.2)[d]	1.2	3.4	100.0	66.7	15.8
Generalized anxiety disorder	15	1.7 (1.5-1.9)[d]	0.7	2.7	86.7	76.9	23.7[d]
Agoraphobia	8	2.0 (1.7-2.3)[d]	1.4	4.6	100.0	62.5	18.5
Social phobia	13	1.3 (1.2-1.4)[d]	1.1	2.0	100.0	46.2	13.8
Specific phobia	12	2.0 (1.9-2.2)[d]	1.3	3.1	100.0	100.0	39.3[d]
Separation anxiety disorder	4	1.6 (1.4-1.8)[d]	1.4	2.0	100.0	75.0	3.8
Posttraumatic stress disorder	14	2.6 (2.2-2.9)[d]	1.3	6.4	100.0	78.6	30.6[d]
Any anxiety disorder	15	1.7 (1.6-1.8)[d]	1.2	3.2	100.0	86.7	41.2[d]
Externalizing disorders							
Attention-deficit/hyperactivity disorder	5	0.6 (0.5-0.8)[d]	0.3	0.6	100.0	20.0	6.9
Conduct disorder	3	0.5 (0.4-0.7)[d]	0.3	0.6	100.0	100.0	29.5[d]
Intermittent explosive disorder	6	0.7 (0.6-0.8)[d]	0.4	0.8	100.0	33.3	5.3
Oppositional defiant disorder	3	0.8 (0.6-1.0)[d]	0.5	0.8	100.0	33.3	18.2[d]
Any externalizing disorder	12	0.7 (0.6-0.8)[d]	0.3	1.4	83.3	40.0	7.2
Substance disorders							
Alcohol abuse	15	0.2 (0.2-0.3)[d]	0.1	0.4	100.0	100.0	128.6[d]
Alcohol dependence	11	0.3 (0.3-0.4)[d]	0.1	0.4	100.0	100.0	95.6[d]
Drug abuse or dependence	5	0.4 (0.3-0.4)[d]	0.1	0.4	100.0	100.0	35.3[d]
Any substance disorder	14	0.3 (0.2-0.3)[d]	0.1	0.4	100.0	100.0	141.6[d]
Any disorder	15	1.1 (1.1-1.2)[d]	0.7	2.2	66.7	70.0	111.5[d]

Abbreviations: CI, confidence interval; F:M, female to male; Max, maximum; Min, minimum; OR, odds ratio; WMH, World Health Organization World Mental Health.

[a] Based on discrete time survival models that used respondent cohort (age at interview), sex, and person-year to predict first onset of each disorder both separately in each country and pooled across all countries.

[b] The number of countries differs across outcomes because not all disorders were assessed in all countries. The pooled results for any mood disorder, any anxiety disorder, any externalizing disorder, and any substance disorder pooled whatever disorders in the relevant category existed in the country across countries though the set sometimes differed across countries.

[c] Based on the pooled cross-national survival model that included an interaction between gender and the dummy variables for country.

[d] Significant at $P \leq .05$, 2-sided.

account for this variation. The best-fitting model was determined as that with the lowest value on the Bayesian information criterion,[40] a standard measure of model fit. Structural equation models are regression models that estimate coefficients simultaneously across a series of equations, some of which can include presumed latent (not directly measured) variables that are assumed to have a prespecified relationship to measured variables, in an effort to maximize the fit between predicted and observed matrices of covariation among the observed variables. In our case, the structural equation models assumed that time-space variation in a latent measure of gender role traditionality predicted time-space variation in the F:M ORs of disorders that were found to have significant time-space variation.

Survival coefficients and their standard errors were estimated using the Taylor series linearization method[41] in the SUDAAN software system.[42] Multivariate tests of the significance of interactions involving gender with person-year, cohort, and country were made with Wald χ^2 tests using Taylor series design-based coefficient variance-covariance matrices. In the case of cross-national models, a single variable that assigned a unique value to each sampling stratum in each country was created, while a second variable that distinguished sampling-error calculation units within each stratum was created. These 2 variables were used as the input to SUDAAN to calculate design-based estimates. Structural equation models were calculated using the Mplus software system.[43] Significance tests of regression coefficients in the structural equation models were estimated using the standard errors generated by Mplus, which assumed that the 58 time-space observations represented a simple random sample from a larger universe of such units. $P \leq .05$ (2-sided) was considered significant.

RESULTS

GENDER DIFFERENCES IN LIFETIME RISKS OF MENTAL DISORDERS

Results are highly consistent across countries in showing that women have a significantly higher lifetime risk of most mood disorders (MDD and dysthymic disorder) and all anxiety disorders than men (**Table 2**). The pooled F:M ORs for these disorders are all statistically significant and range from 1.3 to 2.6. Within-country ORs for these disorders are also consistently greater than 1.0. The one exception to this general pattern is bipolar disor-

Table 3. Interactions of Gender With Cohort in Predicting Lifetime Risk of *DSM-IV* Mental Disorders in the WMH Surveys (N=72 933)[a]

Mental Disorder	No. of Subjects[b]	All-Country OR (95% CI)	Within-Country OR Variation			
			Range		%	
			Min	Max	Dominant Direction	$P \leq .05$ in Dominant Direction
Mood disorders						
Major depressive disorder	15	0.9 (0.8-1.0)[c]	0.6	1.1	73.3	36.4
Dysthymic disorder	8	0.9 (0.7-1.1)	0.2	1.5	50.0	25.0
Bipolar disorder	6	1.1 (0.9-1.3)	0.9	2.1	66.7	0.0
Any mood disorder	15	0.9 (0.8-0.9)[c]	0.6	1.2	60.0	33.3
Anxiety disorders						
Panic disorder	12	1.0 (0.8-1.1)	0.3	3.5	58.3	0.0
Generalized anxiety disorder	15	1.1 (1.0-1.2)	0.4	1.7	60.0	11.1
Agoraphobia	8	1.0 (0.8-1.2)	0.7	1.6	50.0	0.0
Social phobia	12	1.0 (0.9-1.1)	0.7	2.3	58.3	14.3
Specific phobia	12	1.0 (0.9-1.1)	0.7	1.2	50.0	16.7
Separation anxiety disorder	4	0.9 (0.7-1.1)	0.7	1.3	50.0	0.0
Posttraumatic stress disorder	14	1.1 (1.0-1.3)	0.4	2.0	57.1	25.0
Any anxiety disorder	15	1.0 (0.9-1.0)	0.7	1.2	53.3	0.0
Externalizing disorders						
Attention-deficit/hyperactivity disorder	5	1.0 (0.7-1.5)	0.1	5.9	80.0	0.0
Conduct disorder	3	1.4 (0.8-2.2)	0.1	1.5	66.7	100.0
Intermittent explosive disorder	6	1.3 (1.1-1.5)[c]	0.4	1.8	83.3	40.0
Oppositional defiant disorder	3	0.9 (0.5-1.6)	0.7	0.9	100.0	0.0
Any externalizing disorder	12	1.1 (0.8-1.4)	0.2	6.5	58.3	28.6
Substance disorders						
Alcohol abuse	14	1.5 (1.3-1.7)[c]	0.7	4.0	85.7	66.7
Alcohol dependence	10	1.5 (1.2-1.8)[c]	0.8	6.6	90.0	55.6
Drug abuse or dependence	3	1.2 (0.9-1.5)	0.6	1.3	66.7	50.0
Any substance disorder	14	1.4 (1.3-1.7)[c]	0.7	5.0	85.7	66.7
Any disorder	15	1.0 (0.9-1.0)	0.8	1.7	60.0	0.0

Abbreviations: CI, confidence interval; F:M, female to male; Max, maximum; Min, minimum; OR, odds ratio; WMH, World Health Organization World Mental Health.

[a] Based on discrete time survival models that used respondent cohort (age at interview), sex, person-year, and the interaction between sex and cohort to predict first onset of each disorder both separately in each country and pooled across all countries. Each respondent was classified as being in 1 of 4 cohorts based on age at interview (18-34, 35-49, 50-64, and ≥65 years). A continuous variable coded 1 through 4, 1 for the earliest cohorts (ie, aged ≥65 years at interview) and 4 for the most recent cohorts (ie, aged 18-34 years at interview) was used to create the interaction terms, for which the variable was multiplied by a gender variable (0-1) (coded 1 for women and 0 for men). An interaction term with an OR significantly greater than 1.0 means that the F:M OR is significantly higher (ie, higher relative prevalence among women than men) in more recent cohorts, while an OR less than 1.0 means that the F:M OR is significantly lower (ie, lower relative prevalence among women than men) in more recent cohorts.

[b] The number of countries differs across outcomes because not all disorders were assessed in all countries. The pooled results for any mood disorder, any anxiety disorder, any externalizing disorder, and any substance disorder pooled whatever disorders in the relevant category existed in the country across countries, though the set sometimes differed across countries.

[c] Significant at $P \leq .05$, 2-sided.

der, for which the pooled OR is not statistically significant (OR, 0.9). Results are the opposite for most externalizing disorders (attention-deficit/hyperactivity disorder, conduct disorder, and intermittent explosive disorder) and all substance disorders. The pooled F:M ORs for these disorders are all statistically significant and less than 1.0 (range, 0.3-0.8), indicating significantly higher risk among men than women. Within-country ORs for these disorders are also always less than 1.0. Despite their consistent direction, the magnitudes of the ORs vary significantly across countries for many of the disorders studied.

INTERCOHORT VARIATION

Gender differences in lifetime risk for most disorders do not differ significantly across cohorts (**Table 3**). However, there are 3 notable exceptions. The first involves MDD, which has a pooled OR for the gender × cohort interaction across countries of 0.88 (95% confidence interval, 0.82-0.95). This means that the higher odds among women than men are less pronounced in more recent cohorts. This general pattern is found in 11 of the 15 countries. The second exception involves intermittent explosive disorder, in which the pooled gender × cohort OR across countries is 1.26 (95% confidence interval, 1.07-1.48). This means that the higher odds among men than women are less pronounced in more recent cohorts. This general pattern is found in 5 of the 6 countries that assessed intermittent explosive disorder. The third involves substance disorders, largely driven by alcohol disorders. The pooled gender × cohort OR across countries in predicting any substance disorder is 1.45 (95% confidence interval, 1.27-1.66). It is important to remember that any substance disorder is equivalent to either alcohol or drug abuse because, as noted in the "*DSM-IV* Disorders" section, dependence was assessed only among respondents with a history of abuse. This means that the higher odds of a lifetime substance abuse among men than

women are less pronounced in more recent cohorts. This general pattern is found in 12 of the 14 countries in which substance abuse was assessed.

TIME-SPACE VARIATION IN GENDER ROLE TRADITIONALITY AND F:M ORs

The remaining analyses focused on MDD and substance disorders, the disorders associated with significant intercohort variation in the F:M ORs across most countries. (Intermittent explosive disorder was excluded because it was assessed in only 6 WMH countries, yielding too few time-space units for stable analysis.) Before describing the results, it is instructive to examine the raw data on time-space variation in the composite measure of gender role traditionality, created by averaging standardized scores on the 4 indicators (**Table 4**). A generally monotonic decrease in traditionality can be seen across successively more recent cohorts in each country. All but 2 countries (New Zealand and Ukraine) had levels above the mean in gender role traditionality at the time respondents in the oldest cohorts reached early adulthood. All but 2 countries had levels below the mean and the other 2 (Italy and Lebanon) were very close to the mean gender role traditionality in comparison by the time respondents in the most recent cohorts entered early adulthood. Lebanon had by far the highest traditionality score in each cohort, but the other countries with high gender role traditionality scores in the oldest cohorts were all developed countries either in southern Europe (Italy and Spain) or Asia (Japan). Three of these 4 countries (Japan, Lebanon, and Spain) had the most dramatic decreases in gender role traditionality over time, along with Belgium and the Netherlands.

A number of structural equation models were fit to examine the associations of this time-space variation in gender role traditionality with the F:M ORs in MDD and substance disorder. (Detailed results are available from the corresponding author on request.) The final model (**Figure**) defines gender role traditionality as a standardized (to a mean of 0 and variance of 1) latent variable indicated by our 4 observed measures, in which gender role traditionality is a predictor of the F:M ORs of MDD and any substance disorder. The 4 measured indicators are assumed to have effects on these outcomes only through gender role traditionality. High gender role traditionality score is significantly associated with an increase of 0.38 SDs in the F:M OR for MDD (the high prevalence of MDD in women decreases as female gender roles become less traditional) and with a decrease of 0.46 SD in the F:M OR for any substance disorder (ie, women begin to "catch up" to men in their rates of substance disorders as female gender roles become less traditional).

Because all coefficients in the model are standardized (the measures are transformed to have a mean of 0 and a variance of 1), it is necessary to consider the substantive meaning of a standard deviation for each measure to put the results into meaningful terms. Beginning with the gender role traditionality indicators, a 1-SD decrease in gender role traditionality would be equivalent to changing the F:M ratio of labor force participation from the sample-wide mean of 0.85 (women were about 15% less likely than men to be in the labor force) to 1.0 (women and men were equally likely to be in the labor force), changing the F:M ratio of high educational attainment from the mean of 0.84 (women were 16% less likely than men to obtain high education) to 1.03 (rough gender equality), changing the older age at marriage of men than women from the mean of 3.2 years to 2.2 years, and changing the proportion of young women using birth control from the mean of 37.5% to 69.4%.

As noted above, a decrease of 1 SD in gender role traditionality indicators is associated with a decrease in the F:M OR of MDD of 0.38 SD and an increase in the F:M OR of any substance disorder of 0.46 SD. We need to know the means and standard deviations of the outcomes in the unweighted sample of 58 time-space units to make substantive sense of these effect size estimates. The mean values are 2.61 (SD, 1.81) for MDD and 0.17 (SD, 0.14) for any substance disorder. Therefore, changes of 1 SD in the gender role traditionality indicators are associated with a reduction in the F:M OR for MDD from the mean of 2.6 to 1.9 (a reduction of nearly 45% in the elevated F:M OR) and with an increase in the F:M OR of any substance disorder from the mean of 0.17 to 0.25 (a reduction of nearly 30% in the elevated M:F OR).

Although variations due to time and space were combined in the structural equation analysis, it is possible to separate the 2 components by introducing dummy variable controls in the model either for cohort (time) or for country (space) (**Table 5**). When this is done, we see that the association between gen-

Table 4. Distribution of the Standardized Gender Role Traditionality Composite Measure Across the WMH Countries and Cohorts[a]

	Gender Role Traditionality Score by Age Cohort			
Country	18-34 y	35-49 y	50-64 y	≥65 y
Developed countries				
Belgium	−3.7	−2.8	0.9	2.7
France	−3.5	−2.5	−0.7	0.9
Germany	−3.4	−3.4	0.9	1.6
Israel	−2.4	−0.4	−0.3	0.7
Italy	0.2	1.2	2.9	4.8
Japan	−4.4	0.1	−1.2	3.1
The Netherlands	−5.1	−0.2	−1.9	1.7
New Zealand	−2.2	−0.9	−1.3	−1.3
Spain	−4.0	−0.7	3.3	3.4
United States	−3.5	−3.1	−2.3	2.0
Developing countries				
Colombia	−0.5	1.2	2.0	
Lebanon	0.2	4.4	8.3	8.3
Mexico	−0.6	2.0	3.1	
South Africa	−0.8	1.1	1.2	2.2
Ukraine	−2.1	−4.8	0.2	−1.0

Abbreviation: WMH, World Health Organization World Mental Health.
[a]Each indicator was standardized to have a mean of 0 and a variance of 1.0 across the sample of 58 time-space units. Each time-space unit was assigned equal weight in calculating the mean and variance. The composite was then created by summing the 4 standardized indicator scores and then standardizing the sum to have a mean of 0 and a variance of 1.0 across the sample of 58 time-space units.

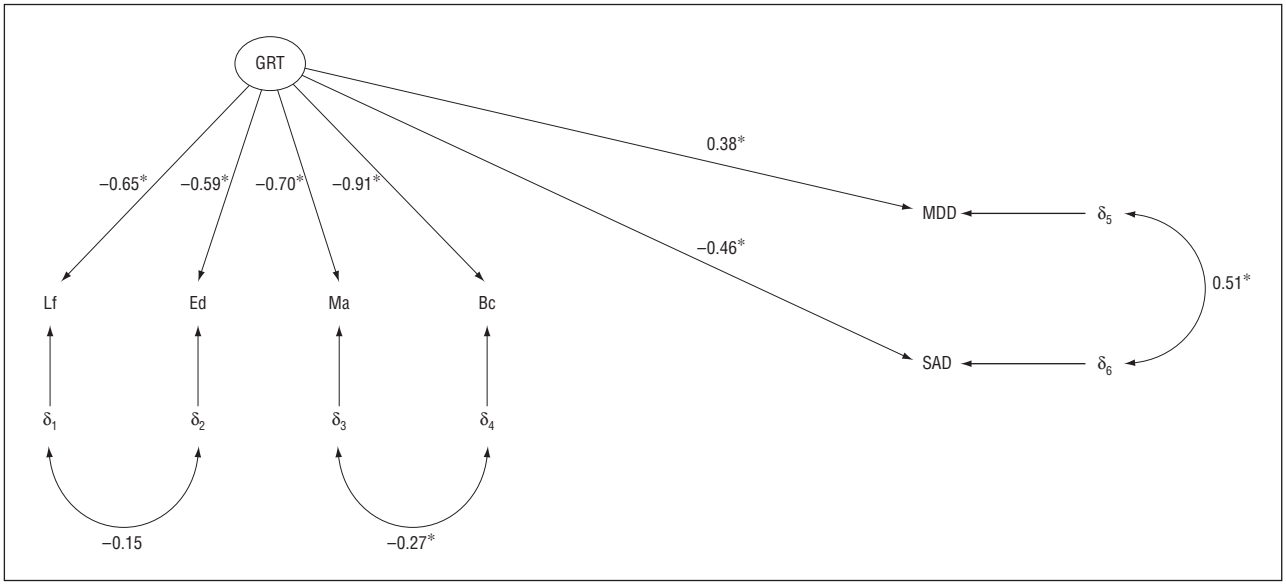

Figure. Standardized parameter estimate of the association between time-space variation in gender role traditionality and the female to male odds ratios of lifetime *DSM-IV* major depressive disorder (MDD) and substance abuse/dependence (SAD) in the 58 WMH cohort by country time-space subsamples. Observed variables and gender role traditionality [GRT] are all standardized to a mean of 0.0 and a standardized deviation of 1.0. Bc indicates percentage of women using birth control before the age of 25 years; Ed, female to male percentage reaching the median education level of the top quartile of earners; Lf, female to male percentage in labor force by age 35 years; Ma, female to male median age at first marriage; *significant at $P \leq .05$, 2-sided.

Table 5. Associations of Gender Role Traditionality With First Onset of Major Depressive Disorder and Any Substance Disorder

	Standardized Regression Coefficient Linking Gender Role Traditionality With Disorder (SE)	
Control	Major Depressive Disorder	Any Substance Disorder
None	0.38 (0.19)[a,b]	−0.46 (0.19)[a,b]
Temporal variation	0.06 (0.20)	−0.44 (0.21)[b]
Spatial variation	0.60 (0.24)[b]	−0.20 (0.18)

Abbreviation: SE, standard error.
[a] Coefficients presented in the Figure.
[b] Significant at $P \leq .05$, 2-sided.

der role traditionality and the F:M OR for MDD is entirely due to between-cohort variation within countries, while the association between gender role traditionality and the F:M OR for any substance disorder is largely due to within-cohort variation across countries. It is consequently only the association involving MDD that involves intercohort changes within countries. We also investigated the possibility that the results are sensitive to extreme values in a small number of countries by replicating the Table 5 results 15 times, each time deleting the data from 1 country. (Detailed results are available from the corresponding author on request.) The only case in which a meaningful change in the model coefficients occurred was when we deleted Lebanon, but even in this case the coefficients remained statistically significant and strong in substantive terms. This result demonstrates that the overall study results are not highly sensitive to individual outlier countries.

COMMENT

Several methodological limitations need to be noted in interpreting the WMH results. First, the response rates were lower in developed than developing countries and might have been related to gender role traditionality, possibly introducing bias into the results. We weighted the data in each country for differential nonresponse by census sociodemographic variables, but there is no guarantee that this corrected for biases introduced by incomplete responses. Second, the surveys also differed across countries in other ways, such as the language in which they were administered and the extent to which each country had a tradition of independent public opinion research, which would have allowed respondents to see the survey as a normal undertaking; this might have affected results. Third, diagnoses were based on fully structured interviews administered by lay interviewers rather than on clinician-administered semi-structured interviews. This limitation is somewhat reduced by the fact that WMH clinical reappraisal studies documented generally good concordance between the diagnoses based on the CIDIs and diagnoses based on blinded semi-structured clinical reappraisal interviews.[29] However, as noted in the section on measures, the diagnoses of externalizing disorders were not validated and might be less accurate than those of other disorders. The substance dependence diagnoses had the additional problem of excluding cases of dependence without a history of abuse. The results concerning the broadly defined measures of substance disorders should consequently be interpreted as applying to abuse rather than to dependence. As noted in the section on measures, however, empirical studies have shown that the number of cases of substance dependence without a history of abuse is small in the United States and that their exclusion does not meaningfully affect

the size of the coefficients between predictors and measures of substance disorders. Fourth, lifetime prevalence and age at onset were assessed with retrospective reports, which could be systematically biased.[44] We used an innovative probing method designed to minimize recall bias,[34] but bias may still have been introduced by age-related gender differences either in memory failure, mental health awareness, or willingness to admit emotional problems to an interviewer. Fifth, the analysis of gender differences in lifetime risk across cohorts could be biased if the increasing attrition with age in older cohorts is differentially related to history of mental disorders among women vs men. Sixth, the indicators of female gender role traditionality were few and might not have captured all the dimensions of female gender roles that are important for explaining secular trends in the F:M ORs. Furthermore, the indicators we used might be related to constructs other than gender role traditionality. However, these concerns are reduced because our composite gender role traditionality measure correlates very strongly with an independent measure of gender empowerment based on objective administrative data assembled by the Global Economic Forum.

Within the context of these limitations, our article is the first to present the results of a quantitative examination of time-space variation in the association between female gender role traditionality and gender differences in mental illness. We found that the frequently observed gender differences in anxiety, mood, externalizing, and substance disorders have remained relatively stable during the more than half century separating respondents in the youngest and oldest WMH cohorts despite the unprecedented changes in female gender roles that occurred during this period. Furthermore, we found that aggregate F:M ORs are relatively consistent across countries despite substantial between-country variation in female gender role traditionality. These patterns argue against the claim that changes in gender roles play an important role in bringing about reductions in gender differences in the lifetime risk of most mental disorders.

The only notable exceptions to this general pattern concern major depression, intermittent explosive disorder, and substance disorders, in which gender differences were found to be significantly smaller in more recent cohorts. As noted in the introduction, evidence consistent with this narrowing has been found in several within-country studies of gender differences in MDD,[9,10,45] substance use disorders,[11] and substance disorders,[46-48] while other studies have documented cross-national variation in gender differences at specific points.[13,22] However, no previous study combined the 2 types of comparisons to study time-space variation while linking independent measures of gender role traditionality to data on variation in gender differences over time or in space.

In the case of MDD, the gender roles hypothesis would interpret our findings as meaning that increases in female opportunities in the domains of employment, birth control, and other indicators of increasing gender role equality promote improvements in female mental health by reducing exposure to stressors that can lead to depression and by increasing access to effective stress-buffering resources.[15,49] However, it is important to acknowledge that we did not directly evaluate the validity of this hypothesis. We documented that gender differences in the risk of MDD onset are significantly narrower for times and places in which the roles of women are more equal to those of men, but we did not measure time-space variation in stress exposure or stress reactivity to see if the latter mediates the predictive effects of gender role traditionality. These more specific analyses go beyond the limits of the WMH data but should be the subject of future studies.

In the case of substance disorders, the gender roles hypothesis takes a somewhat different form by arguing that opportunities for female substance use and attitudes about the appropriateness of female substance use both change as female roles become more similar to male roles, resulting in an increase in female substance use.[14,50,51] Consistent with this hypothesis, cross-national research has documented that current female drinking behaviors are more similar to those in males in countries in which gender roles are more equal,[52,53] though results have not been entirely consistent.[54] However, these previous studies did not examine gender differences in lifetime risk of substance disorders, which means that our results can be seen as building on the findings of these earlier studies.

It is unclear why the WMH findings showed much stronger evidence for temporal than spatial predictive associations in MDD and for spatial than temporal associations in substance disorders. It is also unclear why the narrowing of gender differences in recent cohorts was confined to MDD, intermittent explosive disorder, and substance disorders. It is conceivable that the much younger ages at onset of most other disorders, especially the anxiety disorders and externalizing disorders, than MDD or substance disorders make the former disorders less susceptible than the latter ones to the influences of changes in adult gender roles. That argument does not extend to generalized anxiety disorder, though, which has a similar age at onset distribution to MDD.[55] For this reason, why narrowing occurred for MDD but not generalized anxiety disorder is especially puzzling. Increased understanding of these specifications should be the subject of future theorizing and empirical investigation.

Submitted for Publication: July 28, 2008; final revision received November 4, 2008; accepted November 11, 2008.
Author Affiliations: Medical Research Council Research Unit on Anxiety and Stress Disorder, Cape Town, South Africa (Dr Seedat); Department of Psychological Medicine, Wellington School of Medicine, Otago University, Wellington, New Zealand (Dr Scott); Center for Public Mental Health, Vienna, Austria (Dr Angermeyer); Institute for Social Research, University of Michigan, Ann Arbor (Mss Berglund and Pennell); State University of New York–Stony Brook, Stony Brook (Dr Bromet); Department of Health Sciences, University of Leicester, Leicester, England (Dr Brugha); University Hospital, Gasthuisberg, Leuven, Belgium (Dr Demyttenaere); IRCCS Centro S. Giovanni di Dio Fatebenefratelli, Brescia, Italy (Dr de Girolamo); Sant Joan de Deu-SSM, CIBERSAM, Instituto de Salud Carlos III, Barcelona, Spain (Dr Haro); Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Mr Jin, Ms

Sampson, and Dr Kessler); Institute for Development, Research, Advocacy, and Applied Care (IDRAAC), St George Hospital University Medical Center, Beirut, Lebanon (Dr Karam); University of Paris Descartes, EA 4069; and MGEN Foundation for Public Health, Paris, France (Dr Kovess-Masfety); Ministry of Health, Mental Health Services, Jerusalem, Israel (Dr Levinson); National Institute of Psychiatry, Mexico City, Mexico (Dr Medina Mora); Keio University School of Medicine, Tokyo, Japan (Dr Ono); University Medical Center Groningen, University of Groningen, Groningen, the Netherlands (Dr Ormel); Colegio Mayor de Cundinamarca University, Bogata, Colombia (Dr Posada-Villa); and Harvard University School of Public Health, Boston (Dr Williams).

Correspondence: Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115 (kessler@hcp.med.harvard.edu).

Author Contributions: All authors had access to the data from their own country, but only Ms Berglund, Mr Jin, Ms Sampson, and Dr Kessler had full access to all of the data in the study; Dr Kessler organized responses from all coauthors to the suggestions of reviewers and made final revisions; and he takes responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Seedat, Scott, Angermeyer, Berglund, Brugha, Demyttenaere, de Girolamo, Haro, Jin, Karam, Kovess-Masfety, Levinson, Medina Mora, Ono, Ormel, Pennell, Williams, and Kessler. *Acquisition of data:* Seedat, Scott, Angermeyer, Berglund, Bromet, Demyttenaere, de Girolamo, Jin, Karam, Kovess-Masfety, Levinson, Medina Mora, Ono, Ormel, Pennell, Posada-Villa, Williams, and Kessler. *Analysis and interpretation of data:* Seedat, Scott, Berglund, Sampson, and Kessler. *Drafting of the manuscript:* Seedat, Scott, and Kessler. *Critical revision of the manuscript for important intellectual content:* Seedat, Scott, Angermeyer, Berglund, Bromet, Brugha, Demyttenaere, de Girolamo, Haro, Jin, Karam, Kovess-Masfety, Levinson, Mora, Ono, Ormel, Pennell, Posada-Villa, Sampson, Williams, and Kessler. *Statistical analysis:* Berglund, Jin, and Kessler. *Obtained funding:* Seedat, Angermeyer, Bromet, Brugha, Demyttenaere, de Girolamo, Haro, Karam, Kovess-Masfety, Ono, Ormel, Posada-Villa, Seedat, Williams, and Kessler. *Administrative, technical, or material support:* Seedat, Angermeyer, Berglund, Brugha, de Girolamo, Jin, Karam, Kovess-Masfety, Medina Mora, Ono, Pennell, Posada-Villa, Sampson, Williams, and Kessler. *Study supervision:* Scott, Pennell, Sampson, and Kessler.

Financial Disclosures: Dr Kessler has been a consultant for GlaxoSmithKline Inc, Kaiser Permanente, Pfizer Inc, Sanofi-Aventis, Shire Pharmaceuticals, and Wyeth-Ayerst; has served on advisory boards for Eli Lilly & Company and Wyeth-Ayerst; and has had research support for his epidemiological studies from Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Pharmaceuticals Inc, Pfizer Inc, and Sanofi-Aventis. No other authors have any financial disclosures.

Funding/Support: The surveys included in this article were carried out in conjunction with the World Health Organization WMH Survey Initiative. The WMH staff assisted with instrumentation, fieldwork, and data analysis. These activities were supported by grant R01MH070884 from the US National Institute of Mental Health, the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, grants R13-MH066849, R01-MH069864, and R01 DA016558 from the US Public Health Service, grant FIRCA R03-TW006481 from the Fogarty International Center, the Pan American Health Organization, the Eli Lilly & Company Foundation, Ortho-McNeil Pharmaceutical Inc, GlaxoSmithKline, and Bristol-Myers Squibb. A complete list of WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/. The Mexican National Comorbidity Survey is supported by grant INPRFMDIES 4280 from the National Institute of Psychiatry Ramon de la Fuente and grant CONACyT-G30544- H from the National Council on Science and Technology, with supplemental support from the PanAmerican Health Organization. The Lebanese survey was supported by the Lebanese Ministry of Public Health, the World Health Organization (Lebanon), and unrestricted grants from Janssen Cilag, Eli Lilly, GlaxoSmithKline, Roche, Novartis, Fogerty (grant R03 TW0006481), and anonymous donations. The European Study of the Epidemiology of Mental Disorders is funded by the European Commission (contracts QLG5-1999-01042 and SANCO 2004123), the Piedmont Region (Italy), Fondo de Investigación Sanitaria, Instituto de Salud Carlos III (Spain) (grant FIS 00/0028), Ministerio de Ciencia y Tecnología (Spain) (grant SAF 2000-158-CE), Departament de Salut, Generalitat de Catalunya (Spain), Instituto de Salud Carlos III (grants CIBER CB06/02/0046 and RETICS RD06/0011 REM-TAP), and other local agencies and by an unrestricted educational grant from GlaxoSmithKline. The Colombian National Study of Mental Health is supported by the Ministry of Social Protection. The World Mental Health Japan Survey is supported by the Grant for Research on Psychiatric and Neurological Diseases and Mental Health (grants H13-SHOGAI-023, H14-TOKUBETSU-026, and H16-KOKORO-013) from the Japan Ministry of Health, Labour and Welfare. The New Zealand Mental Health Survey is supported by the New Zealand Ministry of Health, Alcohol Advisory Council, and the Health Research Council. The South Africa Stress and Health Study is supported by grant R01-MH059575 from the US National Institute of Mental Health and National Institute of Drug Abuse, with supplemental funding from the South African Department of Health and the University of Michigan (grant HHSN271200700030C from the National Institutes of Mental Health). The Ukraine Comorbid Mental Disorders During Periods of Social Disruption study is funded by grant RO1-MH61905 from the US National Institute of Mental Health. The US National Comorbidity Survey Replication is supported by grant U01-MH60220 from the National Institute of Mental Health with supplemental support from the National Institute of Drug Abuse, the Substance Abuse and Mental Health Services Administration, grant 044780 from the Robert Wood Johnson Foundation, and the John W. Alden Trust.

Role of the Sponsors: The study's funders had no role in the design or conduct of the study; in the collection, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript.

REFERENCES

1. Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatr Scand*. 2003;108(3):163-174.
2. Pigott TA. Gender differences in the epidemiology and treatment of anxiety disorders. *J Clin Psychiatry*. 1999;60(suppl 18):4-15.
3. Arnold LE. Sex differences in ADHD: conference summary. *J Abnorm Child Psychol*. 1996;24(5):555-569.
4. Brady KT, Randall CL. Gender differences in substance use disorders. *Psychiatr Clin North Am*. 1999;22(2):241-252.
5. Keenan K, Loeber R, Green S. Conduct disorder in girls: a review of the literature. *Clin Child Fam Psychol Rev*. 1999;2(1):3-19.
6. Grigoriadis S, Robinson GE. Gender issues in depression. *Ann Clin Psychiatry*. 2007;19(4):247-255.
7. Lynch WJ, Roth ME, Carroll ME. Biological basis of sex differences in drug abuse: preclinical and clinical studies. *Psychopharmacology (Berl)*. 2002;164(2):121-137.
8. Nolen-Hoeksema S, Hilt L. Possible contributors to the gender differences in alcohol use and problems. *J Gen Psychol*. 2006;133(4):357-374.
9. Joyce PR, Oakley-Browne MA, Wells JE, Bushnell JA, Hornblow AR. Birth cohort trends in major depression: increasing rates and earlier onset in New Zealand. *J Affect Disord*. 1990;18(2):83-89.
10. Wickramaratne PJ, Weissman MM, Leaf PJ, Holford TR. Age, period and cohort effects on the risk of major depression: results from five United States communities. *J Clin Epidemiol*. 1989;42(4):333-343.
11. Bloomfield K, Gmel G, Neve R, Mustonen H. Investigating gender convergence in alcohol consumption in Finland, Germany, The Netherlands, and Switzerland: a repeated survey analysis. *Subst Abus*. 2001;22(1):39-53.
12. McPherson M, Casswell S, Pledger M. Gender convergence in alcohol consumption and related problems: issues and outcomes from comparisons of New Zealand survey data. *Addiction*. 2004;99(6):738-748.
13. Wilsnack RW, Vogeltanz ND, Wilsnack SC, Harris TR, Ahlstrom S, Bondy S, Csemy L, Ferrence R, Ferris J, Fleming J, Graham K, Greenfield T, Guyon L, Haavio-Mannila E, Kellner F, Knibbe R, Kubicka L, Loukomskaia M, Mustonen H, Nadeau L, Narusk A, Neve R, Rahav G, Spak F, Teichman M, Trocki K, Webster I, Weiss S. Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns. *Addiction*. 2000;95(2):251-265.
14. Pape H, Hammer T, Vaglum P. Are "traditional" sex differences less conspicuous in young cannabis users than in other young people? *J Psychoactive Drugs*. 1994;26(3):257-263.
15. Thoits PA. Multiple identities: examining gender and marital status differences in distress. *Am Sociol Rev*. 1986;51:259-272.
16. Wauterickx N, Bracke P. Unipolar depression in the Belgian population: trends and sex differences in an eight-wave sample. *Soc Psychiatry Psychiatr Epidemiol*. 2005;40(9):691-699.
17. Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P, Wells KB, Zaslavsky AM. Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med*. 2005;352(24):2515-2523.
18. Roos E, Lahelma E, Rahkonen O. Work-family conflicts and drinking behaviours among employed women and men. *Drug Alcohol Depend*. 2006;83(1):49-56.
19. Sachs-Ericsson N, Ciarlo JA. Gender, social roles, and mental health: an epidemiological perspective. *Sex Roles*. 2000;43(9-10):605-628.
20. Weich S, Sloggett A, Lewis G. Social roles and gender difference in the prevalence of common mental disorders. *Br J Psychiatry*. 1998;173:489-493.
21. Kessler RC, Üstün TB, eds. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. Cambridge, England: Cambridge University Press; 2008.
22. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H, Yeh EK. Cross-national epidemiology of major depression and bipolar disorder. *JAMA*. 1996;276(4):293-299.
23. World Bank. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2003.
24. Heeringa SG, Wells JE, Hubbard F, Mneimneh Z, Chiu WT, Sampson NA, Berglund PA. Sample Designs and Sampling Procedures. In: Kessler RC, Üstün TB, eds. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. Cambridge, England: Cambridge University Press; 2008:14-32.
25. Pennell BE, Mneimneh Z, Bowers A, Chardoul S, Wells JE, Viana MC, Dinkelmann K, Gebler N, Florescu S, He Y, Huang Y, Tomov T, Vilagut G. Implementation of the World Mental Health Surveys. In: Kessler RC, Üstün TB, eds. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. Cambridge, England: Cambridge University Press; 2008:33-57.
26. Kessler RC, Üstün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):93-121.
27. Harkness J, Pennell BE, Villar A, Gebler N, Aguilar-Gaxiola S, Bilgen I. Translation procedures and translation assessment in the World Mental Health Survey Initiative. In: Kessler RC, Üstün TB, eds. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. Cambridge, England: Cambridge University Press; 2008:91-113.
28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
29. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA, Kessler RC. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *Int J Methods Psychiatr Res*. 2006;15(4):167-180.
30. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, Howes MJ, Jin R, Secnik K, Spencer T, Üstün TB, Walters EE. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med*. 2005;35(2):245-256.
31. Degenhardt L, Bohnert KM, Anthony JC. Case ascertainment of alcohol dependence in general population surveys: "gated" versus "ungated" approaches. *Int J Methods Psychiatr Res*. 2007;16(3):111-123.
32. Degenhardt L, Bohnert KM, Anthony JC. Assessment of cocaine and other drug dependence in the general population: "gated" versus "ungated" approaches. *Drug Alcohol Depend*. 2008;93(3):227-232.
33. Degenhardt L, Cheng H, Anthony JC. Assessing cannabis dependence in community surveys: methodological issues. *Int J Methods Psychiatr Res*. 2007;16(2):43-51.
34. Knauper B, Cannell CF, Schwarz N, Bruce ML, Kessler RC. Improving the accuracy of major depression age of onset reports in the US National Comorbidity Survey. *Int J Methods Psychiatr Res*. 1999;8(1):39-48.
35. Fischer AH, Rodriguez Mosquera PM, van Vianen AE, Manstead AS. Gender and culture differences in emotion. *Emotion*. 2004;4(1):87-94.
36. Schmitt DP. Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating. *Behav Brain Sci*. 2005;28(2):247-311.
37. Wood W, Eagly AH. A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. *Psychol Bull*. 2002;128(5):699-727.
38. Lopez-Claros A, Zahidi S. *Women's Empowerment: Measuring the Global Gender Gap*. Geneva, Switzerland: World Economic Forum; 2005.
39. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc*. 1988;83:414-425.
40. Raftery AE. Bayesian model selection in social research. In: Marsden PV, ed. *Sociological Methodology*. Cambridge, MA: Blackwell Publishers; 1995:111-163.
41. Wolter KM. *Introduction to Variance Estimation*. New York, NY: Springer-Verlag; 1985.
42. *SUDAAN: Professional Software for Survey Data Analysis*. Version 8.0.1. Research Triangle Park, NC: Research Triangle Institute; 2002.
43. Mplus. Version 5. Los Angeles, CA: Muthén & Muthén; 2007.
44. Wilhelm K, Parker G, Geerligs L, Wedgwood L. Women and depression: a 30 year learning curve. *Aust N Z J Psychiatry*. 2008;42(1):3-12.
45. Murphy JM, Laird NM, Monson RR, Sobol AM, Leighton AH. Incidence of depression in the Stirling County Study: historical and comparative perspectives. *Psychol Med*. 2000;30(3):505-514.
46. Grant BF. Prevalence and correlates of drug use and DSM-IV drug dependence in the United States: results of the National Longitudinal Alcohol Epidemiologic Survey. *J Subst Abuse*. 1996;8(2):195-210.
47. Keyes KM, Grant BF, Hasin DS. Evidence for a closing gender gap in alcohol use, abuse, and dependence in the United States population. *Drug Alcohol Depend*. 2008;93(1-2):21-29.
48. Nelson CB, Heath AC, Kessler RC. Temporal progression of alcohol dependence symptoms in the U.S. household population: results from the National Comorbidity Survey. *J Consult Clin Psychol*. 1998;66(3):474-483.
49. Jenkins R. Sex differences in minor psychiatric morbidity. *Psychol Med Monogr Suppl*. 1985;7:1-53.
50. Celentano DD, McQueen DV. Alcohol consumption patterns among women in Baltimore. *J Stud Alcohol*. 1984;45(4):355-358.
51. Parker DA, Harford TC. Gender-role attitudes, job competition and alcohol consumption among women and men. *Alcohol Clin Exp Res*. 1992;16(2):159-165.
52. Rahav G, Wilsnack R, Bloomfield K, Gmel G, Kuntsche S. The influence of societal level factors on men's and women's alcohol consumption and alcohol problems. *Alcohol Alcohol Suppl*. 2006;41(1):i47-i55.
53. Wilsnack SC, Wilsnack RW. International gender and alcohol research: recent findings and future directions. *Alcohol Res Health*. 2002;26(4):245-250.
54. Ahlström S, Bloomfield K, Knibbe R. Gender differences in drinking patterns in nine European countries: descriptive findings. *Subst Abus*. 2001;22(1):69-85.
55. Kessler RC, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Anthony JC, Berglund PA, Chatterji S, de Girolamo G, de Graaf R, Demyttenaere K, Gasquet I, Gluzman S, Gruber MJ, Gureje O, Haro JM, Heeringa S, Karam A, Kawakami N, Lee S, Levinson D, Medina-Mora ME, Oakley-Browne MA, Pennell B-E, Petukhova M, Posada-Villa J, Ruscio A, Stein DJ, Tsang CHA, Üstün TB. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Mental Health Survey Initiative. In: Kessler RC, Üstün TB, eds. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. Cambridge, England: Cambridge University Press; 2008:511-521.